METHODS FOR ROOM ACOUSTIC ANALYSIS USING A MONOPOLE-DIPOLE MICROPHONE ARRAY

J. S. Abel and D. R. Begault*

San Jose State University / Human Information Processing Research Branch, NASA Ames Research Center, MS 262-2, Moffett Field, CA 94035-1000

1. INTRODUCTION

In recent work [2], a microphone array consisting of an omnidirectional microphone and colocated dipole microphones having orthogonally aligned dipole axes was used to examine the directional nature of a room impulse response. The arrival of significant reflections was indicated by peaks in the power of the omnidirectional microphone response; reflection direction of arrival was revealed by comparing zero-lag crosscorrelations between the omnidirectional response and the dipole responses to the omnidirectional response power to estimate arrival direction cosines with respect to the dipole axes.

Ideally, a dipole microphone amplifies an incoming signal according to the cosine of its arrival angle θ with respect to the dipole axis φ_D,

$$a_D(\theta) = cos(\theta - \varphi_D). \tag{1}$$

As a result, crosscorrelations between the omnidirectional microphone response $m_M(t)$ and the dipole responses $m_{Di}(t)$ normalized by the omnidirectional response power, termed *directional fractions,*

$$C_{D_iM}(t) = \sum_{\tau=t-\delta/2}^{t+\delta/2} m_{D_i}(\tau)m_M(\tau) \Big/ \sum_{\tau=t-\delta/2}^{t+\delta/2} m_M(\tau)m_M(\tau), \tag{2}$$

approximate source direction cosines with respect to the dipole axes for the dominate signal present in the analysis window $[t - \delta/2, t + \delta/2]$. In [2], direction of arrival with

*This work is supported by NASA-San Jose State University co-operative agreement NC2-237.

respect to any pair of axes i and j was estimated for each analysis window as the angle of the measured direction cosines vector,

$$\hat{\theta}_E = \arctan(C_{D_iM}, C_{D_jM}). \tag{3}$$

For purposes of room acoustic analysis, however, this processing method has drawbacks which limit its usefulness. In particular, when arrivals are not well separated in time—as is the case for all but the first few specular reflections in a typical room—the direction of arrival estimates produced can be inaccurate. Further, measurements made using a microphone array with three orthogonally oriented dipole elements may lack the spatial information needed to provide psychoacoustically transparent synthesis of a measured room.

In [1], the performance of the direction of arrival estimate (3) was examined in the case of a signal $s(t), t = 0, 1, \ldots, T-1$ arriving at the microphone array in the presence of additive measurement noise with independent identically distributed samples having zero mean and covariance σ^2. The direction of arrival estimate (3) was found to be unbiased with variance given in units of radians squared by

$$\mathrm{Var}\left\{\hat{\theta}_\mathrm{E}\right\} = \frac{1}{\mathrm{T}\cdot\mathrm{SNR}}, \qquad \mathrm{SNR} = \frac{1}{\mathrm{T}}\sum_{t=0}^{\mathrm{T}-1} \mathrm{s}(t)^2/\sigma^2, \tag{4}$$

where T is the signal duration and SNR is the measurement signal-to-noise ratio. Accordingly, for an arrival to be localized to within, for example, 10 degrees on average, it must appear with a total energy roughly 30 dB above the average noise background. In addition, if other reflections are present in the "noise" background, the directional fractions become weighted sums of arrival angle direction cosines, which are dependent on crosscorrelations among the reflection signals. The resulting angle of arrival estimate may bear no resemblance to the arrival angle of any of the reflections present.

In this paper, a maximum likelihood estimator and likelihood ratio detector are presented which can detect reflections and estimate arrival angles for multiple simultaneously arriving reflections. Rather than determining a single direction of arrival for each analysis window, as proposed in [2], the approach presented here forms a function of arrival direction at each point in time. The function is designed to indicate the power impinging the array as a function of direction of arrival, and will peak at the arrival angles of any reflections present. In this way, simultaneously arriving reflections appear as distinct maxima, separated in arrival angle.

We begin by studying the localization information contained in monopole-dipole microphone array room measurements. The information inequality is used to place a lower bound on the variance of arrival angle estimates as a function of array and room characteristics. The use of additional dipole elements is shown to increase angle of arrival estimate accuracy, particularly in the presence of overlapping reflections. Finally, a likelihood function-based room analysis method is described, and an example room measurement presented.

2. LOCALIZATION ACCURACY

Consider a set of J source signals $s_j(t)$ appearing at a monopole-dipole microphone array, the jth signal arriving from direction θ_j. The microphone array consists of K

colocated monopole and dipole elements which record signals $m_k(t)$ via antenna patterns $a_k(\theta)$ in the presence of additive measurement noise $n_k(t)$. The monopole elements pass all signals unchanged, $a_M(\theta_j) = 1$; whereas the dipole elements amplify incoming signals according to the cosine of the arrival angle with respect to the dipole axis φ_D, $a_D(\theta_j) = \cos(\theta_j - \varphi_D)$. The microphone signals are given by

$$m_k(t) = \sum_{j=1}^{J} a_k(\theta_j) \cdot s_j(t) + n_i(t), \qquad t = 0, 1, \ldots, T-1, \tag{5}$$

and may be expressed in matrix form as

$$\boldsymbol{M} = \boldsymbol{S}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{N}, \tag{6}$$

where $\boldsymbol{M}$ is the matrix of microphone responses $\boldsymbol{m}_k = [\, m_k(0) \quad \cdots \quad m_k(T-1)\,]^{\mathsf{T}}$, $\boldsymbol{A}$ is the matrix of array steering columns $\boldsymbol{a}(\theta_j) = [\, a_1(\theta_j) \quad \cdots \quad a_K(\theta_j)\,]^{\mathsf{T}}$,

$$\boldsymbol{M} = [\, \boldsymbol{m}_1 \quad \cdots \quad \boldsymbol{m}_K \,], \qquad \boldsymbol{A} = [\, \boldsymbol{a}(\theta_1) \quad \cdots \quad \boldsymbol{a}(\theta_J)\,], \tag{7}$$

$\boldsymbol{S}$ is the matrix of source signals $\boldsymbol{s}_j = [\, s_j(0) \quad \cdots \quad s_j(T-1)\,]^{\mathsf{T}}$, and $\boldsymbol{N}$ is the matrix of microphone measurement noise samples $\boldsymbol{n}_k = [\, n_k(0) \quad \cdots \quad n_k(T-1)\,]^{\mathsf{T}}$,

$$\boldsymbol{S} = [\, \boldsymbol{s}_1 \quad \cdots \quad \boldsymbol{s}_J \,], \qquad \boldsymbol{N} = [\, \boldsymbol{n}_1 \quad \cdots \quad \boldsymbol{n}_K \,]. \tag{8}$$

The measurement noise is assumed to be zero mean with Gaussian-distributed samples, and be independent and identically distributed microphone to microphone, $\wp(\overrightarrow{\boldsymbol{N}}) = \mathcal{N}(0, \boldsymbol{I} \otimes \boldsymbol{\Sigma})$, where $\overrightarrow{\cdot}$ represents the stack of the columns of its matrix argument, and $\otimes$ is the Kronecker product defined by

$$\boldsymbol{F} \otimes \boldsymbol{G} \stackrel{\text{def}}{=} \begin{bmatrix} f_{11}\boldsymbol{G} & \cdots & f_{1R}\boldsymbol{G} \\ \vdots & & \vdots \\ f_{P1}\boldsymbol{G} & \cdots & f_{PR}\boldsymbol{G} \end{bmatrix}, \tag{9}$$

$\boldsymbol{F}$ being a $P \times R$ matrix with elements f_{pr}. The set of microphone responses is therefore Gaussian-distributed with a mean dependent on the unknown signals $\boldsymbol{S}$ and directions of arrival $\boldsymbol{\theta} = [\, \theta_1 \quad \cdots \quad \theta_J \,]^{\mathsf{T}}$,

$$\wp(\overrightarrow{\boldsymbol{M}}; \boldsymbol{\theta}, \boldsymbol{S}) = \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{S}), \boldsymbol{I} \otimes \boldsymbol{\Sigma}), \qquad \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{S}) = \sum_{j=1}^{J} \boldsymbol{a}(\theta_j) \otimes \boldsymbol{s}_j. \tag{10}$$

Information Inequality

Of interest is the ability of the microphone array to determine the arrival direction of each of the signals present. In the absence of a particular processing method, the *information inequality*—often referred to as the Cramer-Rao bound [3, pp. 123–129]—can provide a gauge of the reflection information contained in array measurements by placing an upper limit on the accuracy with which directions of arrival may be determined using microphone array measurements.

For the problem at hand, the information inequality states that the variance in estimating the unknown directions of arrival θ and reflection signals S based on microphone responses M has a lower limit:

$$\operatorname{Var}\left\{\begin{bmatrix}\hat{\theta}\\ \hat{\vec{S}}\end{bmatrix}\right\} \geq \mathcal{F}_{\theta,S}^{-1}, \tag{11}$$

where the direction of arrival and arrival signal estimates $\hat{\theta}$ and $\hat{S}$ are assumed to be unbiased, and $\mathcal{F}_{\theta,S}$ is the *Fisher information matrix*, given by

$$\mathcal{F}_{\theta,S} = \mathrm{E}\left\{\frac{\partial \ln \wp(\vec{M};\theta,S)}{\partial\begin{bmatrix}\theta & \vec{S}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}}\cdot\frac{\partial \ln \wp(\vec{M};\theta,S)}{\partial\begin{bmatrix}\theta & \vec{S}^{\mathsf{T}}\end{bmatrix}}\right\}, \tag{12}$$

where $\mathrm{E}\{\cdot\}$ represents expected value.

Angle of Arrival Estimate Accuracy

In view of (10) and (11), the covariance of arrival angle estimates is bounded below,

$$\operatorname{Var}\left\{\hat{\theta}\right\} \geq \Delta^{-1} + \Delta^{-1}\Upsilon^{\mathsf{T}}(\Gamma - \Upsilon\Delta^{-1}\Upsilon^{\mathsf{T}})^{-1}\Upsilon\Delta^{-1}, \tag{13}$$

where Γ is the Kronecker product of the steering matrix inner product and the measurement noise covariance inverse, and Δ is the element-by-element product of the array steering matrix derivative inner product and source signal matrix inner product,

$$\Gamma = (A^{\mathsf{T}}A)\otimes\Sigma^{-1}, \quad \Delta = (D^{\mathsf{T}}D)\circ(S^{\mathsf{T}}\Sigma^{-1}S), \quad \Upsilon = (\Gamma^{\mathsf{T}}D)\odot(\Sigma^{-1}S), \tag{14}$$

where $\odot$ represents the Khatri-Rao product defined for matrices F and G, each having J columns $f_{\cdot j}$ and $g_{\cdot j}$, $j = 1,\cdots,J$, by

$$F\odot G \stackrel{\text{def}}{=} [f_{\cdot 1}\otimes g_{\cdot 1} \quad \cdots \quad f_{\cdot J}\otimes g_{\cdot J}], \tag{15}$$

and where the steering matrix derivative is given by $D = \begin{bmatrix}\frac{\partial a(\theta_1)}{\partial\theta_1} & \cdots & \frac{\partial a(\theta_J)}{\partial\theta_J}\end{bmatrix}$.

To explore the bound variance behavior as a function of source and array particulars, it is useful to first consider the case of a single arrival,

$$\operatorname{Var}\left\{\hat{\theta}\right\} \geq \frac{a^{\mathsf{T}}a/s^{\mathsf{T}}\Sigma^{-1}s}{\left(\frac{\partial a}{\partial\theta}^{\mathsf{T}}\frac{\partial a}{\partial\theta}\right)a^{\mathsf{T}}a - \left(\frac{\partial a}{\partial\theta}^{\mathsf{T}}a\right)^2}. \tag{16}$$

The bound variance is the inverse product of two terms: $s^{\mathsf{T}}\Sigma^{-1}s$ which is the product of signal duration and signal-to-noise ratio, and a geometric factor roughly equal to one plus half of the number of dipole elements. In other words, more powerful signals may be more accurately localized than weaker ones. In addition, accuracy may be increased by including measurements from additional dipole elements.

It should be pointed out that the arrival direction bound covariance is independent of signal direction of arrival when the dipole axes are uniformly spaced on $[0,\pi)$. However, the dipoles receiving the least signal energy provide the most direction of arrival information, as their antenna patterns are the most sensitive to direction of arrival changes.

Similar results hold in the case of multiple sources. When Γ is nonsingular (i.e., when A has full column rank) the bound (13) can be written as

$$\mathrm{Var}\left\{\hat{\theta}\right\} \geq \left[(D^\top P_A^\perp D) \circ (S^\top \Sigma^{-1} S)\right]^{-1}, \quad P_A^\perp = I - A(A^\top A)^{-1} A^\top, \tag{17}$$

where $P_A^\perp$ is a projection orthogonal to the columns of the array steering matrix A. Again, the bound variance is the inverse of a product of two terms, $S^\top \Sigma^{-1} S$ which is proportional the measurement duration, signal-to-noise ratio product, and $D^\top P_A^\perp D$ which is roughly one plus half of the number of dipole elements used.

When the arrival signals are uncorrelated, the signal inner product $S^\top \Sigma^{-1} S$ becomes diagonal, and the direction of arrival bound variances approximate the corresponding single arrival bound variances. However, if the arrival signals are highly correlated, nearby arrivals with similar steering columns $a(\theta_j)$ will have bound variances which are noticeably larger than their corresponding single arrival variances. In addition, the number of highly correlated arrivals which may be simultaneously resolved is limited to the rank of $D^\top P_A^\perp D$—one plus the number of dimensions for a monopole-dipole array.

3. ARRIVAL DETECTION AND ESTIMATION

The following presents an alternative detection-estimation scheme to (2), (3). In the additive Gaussian measurement noise case considered here, comparing the likelihood ratio to a threshold is known to provide the optimal tradeoff between arrival probability of detection and probability of false alarm. Further, in the limit of large Fisher information (that is, in the limit of small estimate errors) the so-called *maximum likelihood estimator* is unbiased with variance achieving the information inequality bound [3, §6.2].

Given microphone array measurements M, the maximum likelihood arrival direction and signal estimates $\hat{\theta}_{ML}$ and $\hat{S}_{ML}$ are those maximizing the loglikelihood,

$$(\hat{\theta}_{\mathrm{ML}}, \hat{S}_{\mathrm{ML}}) = \arg\{\max_{\theta, S} \ell(M; \theta, S)\}, \tag{18}$$

$$\ell(M; \theta, S) = -\frac{1}{2}[\mu(\theta, S) - \vec{M}]^\top \left(I \otimes \Sigma^{-1}\right) [\mu(\theta, S) - \vec{M}]^\top. \tag{19}$$

The loglikelihood function may be written in terms of the arrival directions only by noting that the set of arrival signals $\hat{S}(\theta) = MA(A^\top A)^{-1}$ maximizes the loglikelihood function for any given arrival direction set θ. We have

$$\ell(M; \theta) = -\frac{1}{2}\mathrm{tr}\left\{P_A^\perp M^\top \Sigma^{-1} M P_A^\perp\right\}. \tag{20}$$

Rather than optimize the loglikelihood $\ell(M; \theta)$ directly by searching over the set of arrival directions θ, here we compare at each arrival direction θ the loglikelihood assuming J arrivals are present, one appearing at θ, with the loglikelihood assuming $J-1$ arrivals $\tilde{\theta}$ are present, none appearing at θ:

$$\lambda(M; \theta) = \min_{\tilde{\theta}} \frac{a(\theta)^\top M^\top \Sigma^{-1} M a(\theta)}{2a(\theta)^\top a(\theta)} - \ell(M; \tilde{\theta}). \tag{21}$$

The resulting loglikelihood ratio indicates received power as a function of arrival direction, and will peak at signal arrival directions. The optimization over $\tilde{\theta}$ is difficult, and

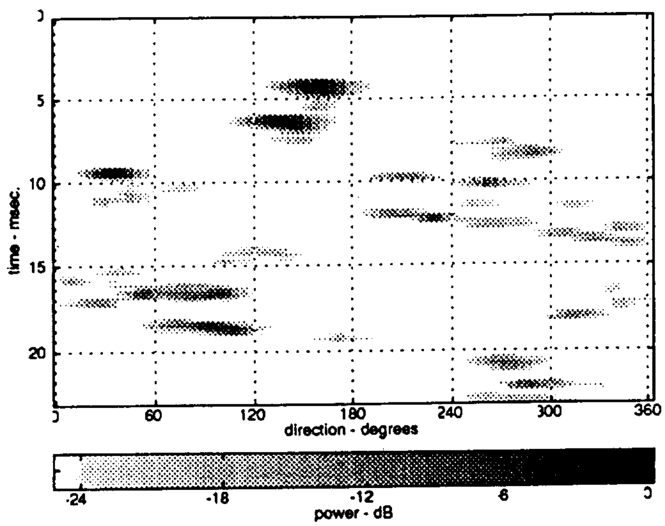

Figure 1: Normalized Likelihood Approximation Example.

$\lambda(M;\theta)$ is approximated as the power in the weighted sum of microphone signals, where the weighting is chosen to minimize the output power while passing unchanged signals arriving from the "look direction" θ.

$$\lambda(M;\theta) = \min_{\gamma} \gamma^{\top} M^{\top} \Sigma^{-1} M \gamma \quad \text{s.t. } a(\theta)^{\top}\gamma = 1. \tag{22}$$

We conclude with an example measurement. A condensor microphone having a rotatable capsule (Neumann USM 69i) was placed in a long narrow hallway along with a sound source (Bose AM-III). The microphone and sound source were placed in the plane perpendicular to the long axis of the hallway, ensuring that all arrivals during the measurement period were also from that plane. Impulse responses were measured using Golay code pairs with the microphone set to receive signals omnidirectionally and via a dipole pattern at a number of dipole axis orientations.

The loglikelihood ratio approximation (22) derived from the measured impulse responses appears in Figure 1. All measured loglikelihood ratio maxima correspond to the arrival times and arrival directions predicted by a ray tracing of the room geometry. In results not shown here, the angular extent of the loglikelihood ratio maxima were seen to decrease with an increasing number of dipole elements. In addition, the likelihood ratio approximation was seen to place distinct maxima at the arrival angles of simultaneously arriving reflections.

References

[1] Begault, D. R., and Abel, J. S., "Studying room acoustics using a monopole-dipole microphone array," in *Proc. 16th ICA/135th ASA*, Paper 1pAAc11, Seattle, June 1998; *J. Acoust. Soc. Am.*, Vol. 103, No. 5, Pt. 2, page 2789, May 1998.

[2] Essert, R., "Measurement of Spatial Impulse Responses with a Soundfield Microphone," in *Paper 5pAA3, Third Joint Meeting of the Acoust. Soc. of Japan*, 1996.

[3] Lehmann, E. L., *Theory of Point Estimation*, J. Wiley and Sons: New York, 1983.